Organization of restriction-modification systems

Geoffrey G.Wilson
New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915, USA

Received February 13, 1991; Revised and Accepted April 10, 1991

ABSTRACT

The genes for over 100 restriction-modification systems have now been cloned, and approximately one-half have been sequenced. Despite their similar function, they are exceedingly heterogeneous. The heterogeneity is evident at three levels: in the gene arrangements; in the enzyme compositions; and in the protein sequences. This paper summarizes the main features of the R-M systems that have been cloned.

INTRODUCTION

Restriction enzymes are traditionally used for cloning genes. In recent years, the genes for many restriction enzymes, themselves, have been cloned. One of the driving forces behind this is commercial: a wish to overproduce the enzymes. The other is academic: a wish to understand how the enzymes recognize DNA. Because restriction enzymes are so specific, and because their specificities are so varied, they offer exceptional opportunities for investigating protein-DNA interactions. Restriction enzymes occur mainly in bacteria, and they are usually accompanied by a modification enzyme of identical specificity; together, the two activities form a restriction-modification (R-M) system—roughly the prokaryotic equivalent of an immune system. Since modification enzymes are as varied, and as specific as, restriction enzymes, there is much interest in studying these enzymes, too.

More than one hundred R-M systems have now been cloned; many have been sequenced. To keep track of what has been accomplished, it would be helpful for workers to have a summary of which genes have been cloned, and the extents to which they have been characterized. The purpose of this paper is to provide that summary. The text of the paper summarizes the properties of R-M systems in general; the tables summarize the main features of the cloned genes in particular. The paper brings up to date an earlier survey (1). For specialized reviews of restriction and modification see refs. 2–8.

RESULTS

Occurrence and function of R-M systems

Restriction-modification systems occur in microorganisms, mainly in bacteria (6). Some temperate bacterial viruses carry R-M systems (9), as do virulent viruses of the unicellular alga, *Chlorella* (10). They have been found in roughly one-quarter of the bacteria examined to date; the remaining three-quarters might lack R-M systems, or they might possess them in forms that have eluded detection. Among the bacteria that demonstrably possess R-M systems, approximately one-half have multiple systems: usually two or three, but sometimes more (11).

Restriction-modification systems protect cells from DNA infections, particularly viral infections. This is probably their sole function. The endonucleases digest foreign DNA that enters the cell, thereby protecting the cell from genetic subversion. The methyltransferases modify the cell's own DNA, thereby protecting it from similar digestion. For practical reasons, few of the several thousand known systems have actually been tested in situ for their ability to restrict viruses, but it is assumed that they all do.

Cloning restriction and modification genes

A striking number of restriction and modification genes have been cloned in recent years. This is largely due to the adoption of a selection that exploits the ability of cloned methyltransferase genes to modify the vector into which they are ligated. Modification enables the plasmid to survive digestion by the corresponding restriction enzyme, and to be recovered following transformation of the digest back into cells. See ref. 12 for a discussion of the technique.

The application of the method at first led to the cloning of only methyltransferase genes (13–16). As it became apparent that *R* and *M* genes are usually linked, and as efforts were made to avoid separating them during cloning, recombinants carrying both genes were recovered (17,18). Over 80 complete R-M systems have now been cloned by this procedure.

Characteristics of R-M systems

Several kinds of R-M systems have been discovered. They appear to do equivalent biological jobs, but in different ways. The differences concern enzyme composition and co-factors; recognition sequence symmetry; and cleavage characteristics (19). Regardless of the kind of system, however, cleavage requires at least Mg^{2+} or a comparable cation; modification requires at least *S*-adenosylmethionine (AdoMet), and affects a single nt in each strand of the recognition sequence. The genes for restriction and modification enzymes appear always to be closely linked.

Type I. Type I systems are complex. They consist of three proteins, R, M, and S, which form an enzyme that restricts and modifies (2,3,20). Cleavage requires AdoMet and ATP;

modification requires AdoMet. Cleavage occurs at considerable, and variable, distances from the recognition sequence (21). The recognition sequences are asymmetric and bipartite; they comprise two sub-sequences, three and four bp in length, separated by six to eight non-specific bp. The S subunit determines specificity for both restriction and modification. The *M* and *S* genes are transcribed as a single operon, and the *R* gene is transcribed separately (Table 1) (22–24).

Type II systems. Type II systems are the simplest and the most common (6). The endonucleases and methyltransferases are separate proteins. The recognition sequences are essentially symmetric. They comprise four to eight specific nt, but they may include additional nt in the form of nonspecific interruptions. Cleavage occurs symmetrically within the sequences. The endonucleases are believed to act mainly as homodimers, the methyltransferases as monomers. The *R* and *M* genes occur in all linkage configurations (25,26). Most often, the genes are aligned; sometimes the *R* gene comes first; at other times the *M* gene comes first. In several systems the genes have opposite orientations; some diverge, others converge (Table 2).

Type IIs R-M systems. Type IIs recognition sequences are asymmetric, uninterrupted, four to seven nt in length. The endonucleases cleave at a defined distance—up to 20 nt—to one side of the sequence. The endonucleases are larger than type II endonucleases, and probably act as monomers (Table 3). Modification is sometimes carried out by two methyltransferases, one for each strand (27–31). In some systems, different nt become methylated on each strand (32). See ref. 8 for a review of type IIs systems.

A few type IIs systems are irregular. *Eco*57I comprises a fused endonuclease-methyltransferase (RM), and a separate methyltransferase (33). The former cleaves outside the recognition sequence, and methylates the sequence on just one strand; the methyltransferase methylates the sequence on both strands. Cleavage is stimulated by AdoMet. *Gsu*I might be similar (33). *Bcg*I is another exception: the endonuclease requires AdoMet, and cleaves outside the recognition sequence, but on both sides. Cleavage excises a 34-bp fragment that contains the recognition sequence (34). *Bst*4.4I might be similar to *Bcg*I (35).

Type III R-M systems. Type III R and M proteins form a complex that restricts and modifies. The M protein also methylates on its own (9). Cleavage requires ATP, and is stimulated by AdoMet. The recognition sequences are asymmetric, uninterrupted, and five to six nt in length (Table 4). Cleavage occurs approximately 25 nt to one side of the sequence. Only one strand of the recognition sequence becomes methylated, in apparent violation of the rule that both strands must be methylated to preserve modification during replication. However, cleavage takes place only when two unmodified sites are present in the DNA, in opposite orientations. Since one site or the other remains modified after passage of the replication fork, modification is preserved during replication (36).

Other types of systems. A number of methyltransferases occur separately, unaccompanied by endonucleases (table 5). Some function in modification (37,38), others are associated with mismatch repair (39,40). Solitary restriction functions also occur. In a few cases the activities are simple endonucleases that require methylated substrates for cleavage e.g. *Dpn*I (41). In other cases, the activities are more complicated (42–48).

Amino acid sequence comparisons

Restriction enzymes vs. modification enzymes. Over fifty type II R-M systems have been sequenced. No similarities have been seen between endonucleases and methyltransferases (49). Some similarities might be expected between companion enzymes since they recognize identical DNA sequences. The lack of similarity suggests that restriction and modification enzymes are unrelated, and that they recognize their targets by different strategies.

Restriction enzymes. Fifty four type II endonucleases have been sequenced. Apart from certain pairs of isoschizomers, the enzymes are dissimilar (49). This suggests that they arose independently during evolution, and not from a common ancestor by divergence of its target recognition domain (TRD). Isoschizomers that cleave the same sequence at the same position ('homoschizomers'?) are sometimes exceptions: *Eco*RI and *Rsr*I (G'AATTC) are closely similar, and probably diverged from a common ancestor (50). Not all homoschizomers are homologous, however: *Hae*III and *Ngo*PII (GG'CC) are entirely dissimilar (51,128). Isoschizomers that cleave the same sequence at different positions ('heteroschizomers'? (7)), for example *Sma*I (CCC'GGG) and *Xma*I (C'CCGGG), are also dissimilar (53,54). Since no common sequence motifs have been discerned among endonucleases, they cannot be recognized as such by inspection of their amino acid sequences.

Modification enzymes. In contrast to the endonucleases, extensive similarities occur among the methyltransferases. Approximately ninety have been sequenced, and seven, or so, architectural classes have been distinguished (55). One class comprises enzymes that form 5-methylcytosine in DNA (m^5C-MTases). Members of this group possess ten, or so, common aa sequence motifs (56). Towards the CO_2H-terminus of these enzymes is a 'variable region' that is believed to form the TRD (57–59). The remaining classes comprise enzymes that form *N*4-methylcytosine (m^4C-MTases), and *N*6-methyladenine (m^6A-MTases). The m^4C-MTases and m^6A-MTases are quite similar, suggesting a common mechanism for methylating the exocyclic amino group of adenine and cytosine (55,60). The enzymes share two principal common sequence motifs. Surprisingly, the order of the motifs differs between certain of the classes (55,61).

Specificity proteins. The specificity of type I enzymes is a function of the S protein (2). S proteins contain two separate TRDs, one for each part of the recognition sequence (62,63). The TRD nearest the NH_2-terminus recognizes the 5′ part of the sequence, and the TRD nearest the CO_2H-terminus recognizes the 3′ part of the sequence (64,65). Between the TRDs is a section that spans the interval between the sequences. Crossovers within this section have generated new R-M system specificities (24,66).

Type I R-M systems

System[a]	Genes[b]	Specificity[c]	Gene organization[d] R / M & S	Refs[e]
CfrA *Citrobacter freundii*	*RMS*	GCAN$_8$GTGG	578	23 65
EcoA *Escherichia coli 15T⁻*	*RMS*	GAGN$_7$GTCA	489 589 m^6A	64 70 71
EcoB *Escherichia coli B*	*MS*	TGAN$_8$TGCT	529 474 m^6A	62 72
EcoD *Escherichia coli E166*	*MS*	TTAN$_7$GTCY	444	62
EcoDXXI *Escherichia coli [pDXX1]*	*RMS*	TCAN$_7$ATTC		73
EcoE *Escherichia coli A58*	*RMS*	GAGN$_7$ATGC	490 594 m^6A	64 70 71
EcoK *Escherichia coli K-12*	*RMS*	AACN$_6$GTGC	1090 529 464 m^6A	13 20 22 62
EcoR124 *Escherichia coli [R124]*	*RMS*	GAAN$_6$RTCG	406 520 m^6A	24
EcoR124/3[1] *Escherichia coli [R124/3]*	*RMS*	GAAN$_7$RTCG	1033 410 520 m^6A	24

[1]*Eco*R124/3 is a spontaneous mutant of *Eco*R124. The S subunit of *Eco*R124/3 contains an extra four amino acids in the part of the protein that separates the proximal and distal sequence-specificity domains. The extra amino acids increase, by one nt, the non-specific interval between the 5' and 3' parts of the recognition sequence (24). The M subunits of *Eco*R124 and *Eco*R124/3 are identical; the R subunits are assumed to be identical, also.

Type I R-M systems

System[a]	Genes[b]	Specificity[c]	Gene organization[d]	Refs[e]
			R *M & S*	
StySB *Salmonella typhimurium LT2*	*RMS*	GAGN$_6$RTAYG	529 469 m^6A	66 72 74
StySJ[2] *Salmonella*	*RMS*	GAGN$_6$GTRC	459 m^6A	66
StySP *Salmonella potsdam*	*RMS*	AACN$_6$GTRC	529 463 m^6A	63 72 74
StySQ *Salmonella*	*RMS*	AACN$_6$RTAYG	473 m^6A	63

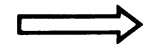
[2] *Sty*SJ and *Sty*SQ are recombinants between the *Sty*SB and *Sty*SP systems. Genetic crossing over between the S genes of *Sty*SB and *Sty*SP, in the interval between the proximal and distal specificity domains, resulted in hybrid S polypeptides that possess the 5' sequence specificity of one parent and the 3' sequence specificity of the other (66,74,75)

Type-II R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
AatII					
Acetobacter aceti	*RM*	GACGT'C			76
AccI			366 541		
Acinetobacter calcoaceticus	*RM*	GT'MKAC		$m^6A\gamma$	12
AflII					
Anabaena aquae-aquae	*RM*	C'TTAAG		m^6A	12
AflIII					
Anabaena flos-aquae	*RM*	A'CRYGT			54
AluI					12
Arthrobacter luteus	*M*	AG'CT		m^5C	77 78
AquI[3]				248+139	
Agmenellum quaduplicatum	*M*	C'YCGRG		m^5C	79
AseI					
Aquaspirillum serpens	*RM*	AT'TAAT			80
AseII					
Aquaspirillum serpens	*RM*	CC'SGG			80
AvaI					
Anabaena variabilis	*RM*	C'YCGRG			12
AvaII					
Anabaena variabilis	*RM*	G'GWCC			12

[3] M·*Aqu*I is unusual in that it comprises two polypeptides, encoded by adjacent ORFs (79). The junction between the ORFs occurs in the variable region of the MTase. The conserved motifs characteristic of the amino-termini of m^5C-MTases occur in the proximal polypeptide, and the conserved motifs characteristic of the carboxy-termini of m^5C-MTases occur in the distal polypeptide.

Type-II R-M systems

System	Genes	Specificity	Gene organization (*R* / *M*)	Refs
AvrI				
Anabaena variabilis UW	*RM*	CYCGRG		54
BalI				
Brevibacterium albidum	*M*	TGG'CCA	m^5C	12, 81
BamHI[4]				
Bacillus amyloliquefaciens	*RM*	G'GATCC	213 423  $m^4C\beta$	82, 85, 86, 87
BanI				
Bacillus aneurinolyticus	*RM*	G'GYRCC	354 428  m^5C	12, 88
BanII				
Bacillus aneurinolyticus	*M*	GRGCY'C		12
BcnI				
Bacillus centrosporus	*RM*	CC'SGG	211  $m^4C\alpha$	15, 89, 90
BepI				
Brevibacterium epidermis	*M*	CG'CG	 m^5C	91
BfaI				
Bacteroides fragilis	*M*	CTAG		54
BglI				
Bacillus globigii	*RM*	GCCN$_4$'NGGC	348  $m^4C\beta$	12, 54, 87
BglII				
Bacillus globigii	*RM*	A'GATCT		92

[4]Immediately preceding the *bamHIR* gene is an ORF (*C*) that regulates expression of the *R* and *M* genes (82,83). Comparable ORFs occur in other systems, especially those in which the gene orientations differ (84).

Type-II R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
***BsaA*I** *Bacillus Stearothermophilus*	*RM*	YAC'GTR			34
***BspE*I** *Bacillus sp.*	*RM*	T'CCGGA			76
***Bsp*HI** *Bacillus sp.*	*RM*	T'CATGA			80
***Bsp*RI** *Bacillus sphaericus R*	*M*	GG'CC		424 ⇨ m^5C	14 93
***Bsp*6I** *Bacillus sphaericus*	*RM*	GC'N_2GC		220 315 ⇨ m^5C	87
***Bsp*50I** *Bacillus sp.*	*M*	CG'CG			94
***Bss*HII** *Bacillus stearothermophilus H3*	*M*	G'CGCGC			76
***Bst*VI** *Bacillus stearothermophilus V*	*RM*	CTCGAG			95
***Bst*XI** *Bacillus stearothermophilus X1*	*RM*	CCAN_5'NTGG			76
***Bst*YI** *Bacillus stearothermophilus Y406*	*M*	R'GATCY			96
***Bsu*BI** *Bacillus subtilis*	*RM*	CTGCAG		311 501 ⇦ $m^6A\gamma$	97

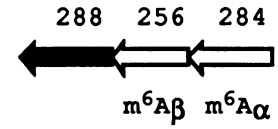
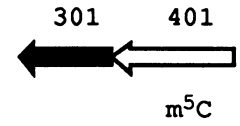
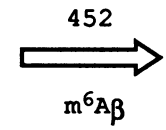
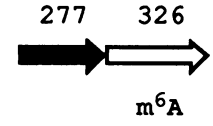
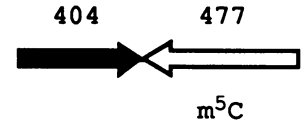
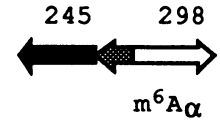
Type-II R-M systems

System	Genes	Specificity	Gene organization (R / M)	Refs
BsuFI *Bacillus subtilis*	*RM*	C'CGG	395 409 → m^5C	98 99
BsuRI[5] *Bacillus subtilis R*	*RM*	GG'CC	576 436 → m^5C	17 100
Bsu15I *Bacillus subtilis*	*RM*	AT'CGAT	m^6A	77
CfrI *Citrobacter freundii RFL2*	*RM*	Y'GGCCR	m^5C	101
Cfr9I *Citrobacter freundii RFL9*	*RM*	C'CCGGG	330 300 → $m^4C\beta$	55 87 90
Cfr10I *Citrobacter freundii RFL10*	*RM*	R'CCGGY	343 → m^5C	90 102
CviJI[6] *Chlorella virus IL-3A*	*M*	RG'CY	367 → m^5C	103
CviRI *Chlorella virus XZ-6E*	*M*	TG'CA	379 → $m^6A\gamma$	105
DdeI *Desulfovibrio desulfuricans*	*RM*	C'TNAG	240 415 → m^5C	106 107

[5]*Bsu*RI is unusual: the *R* gene is large (R·*Bsu*RI is believed to act as a monomer); the *R* and *M* genes are far apart; and, an ORF (unknown function) occurs between them.

[6]The CviJI methyltransferase is less specific than the endonuclease; it methylates RGCB, and possibly VGCB (103,104).

Type-II R-M systems

System	Genes	Specificity	Gene organization (R / M)	Refs
DpnII[7] *Streptococcus pneumoniae*	*RMM*	'GATC		108 109 110 111
DraI *Deinococcus radiophilus*	*RM*	TTT'AAA		112
DraII *Deinococcus radiophilus*	*M*	RG'GNCCY		112
EagI *Enterobacter agglomerans*	*RM*	C'GGCCG		113 114
EcaI *Enterobacter cloacae DSM 30056*	*M*	G'GTNACC		115
EcoRI *Escherichia coli RY13*	*RM*	G'AATTC		116 117
EcoRII *Escherichia coli [N3]*	*RM*	'CCWGG		118 119 120 121
EcoRV *Escherichia coli J62[pLG74]*	*RM*	GAT'ATC		122
Eco47I *Escherichia coli RFL47*	*R*	G'GWCC		102

[7] *Dpn*II codes for two m6A-MTases 108. The distal MTase, M·*Dpn*II, is active only on ds DNA, whereas the proximal MTase, M·*Dpn*A, is active on both ds and ss DNA (109).

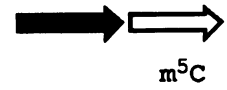
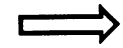
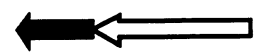
Type-II R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
Eco47II *Escherichia coli RFL47*	*RM*	GGNCC			102
Eco47III *Escherichia coli RFL47*	*M*	AGC'GCT			102
Eco56I *Escherichia coli RFL56*	*M*	G'CCGGC			87
Eco64I *Escherichia coli RFL64*	*RM*	G'GYRCC			123
Eco72I *Escherichia coli RFL72*	*RM*	CAC'GTG	211 372	m^5C	90 123
Eco88I *Escherichia coli RFL88*	*RM*	C'YCGRG			87
Eco98I *Escherichia coli RFL98*	*RM*	A'AGCTT			124
Eco105I *Escherichia coli RFL105*	*RM*	TAC'GTA			124
Eco147I *Escherichia coli RFL147*	*RM*	AGG'CCT			87
FnuDI *Fusobacterium nucleatum D*	*RM*	GG'CC	284 344	m^5C	125 126
FnuDII *Fusobacterium nucleatum D*	*M*	CG'CG		m^5C	12
FnuDIII *Fusobacterium nucleatum D*	*M*	GCG'C			12

Type-II R-M systems

System	Genes	Specificity	Gene organization R / M	Refs
***Fsp*I** *Fischerella sp.*	*RM*	TGC'GCA		81
***Hae*II** *Haemophilus aegyptius*	*RM*	RGCGC'Y	m^5C	127
***Hae*III** *Haemophilus aegyptius*	*RM*	GG'$\underline{C}$C	317 330 m^5C	127 128
***Hgi*AI** *Herpetosiphon giganteus HP1023*	*RM*	GWGCW'C		12
***Hgi*BI** *Herpetosiphon giganteus Hpg5*	*RM*	G'GWCC	274 437 m^5C	129
***Hgi*CI** *Herpetosiphon giganteus Hpg9*	*RM*	G'GYRCC	345 420 m^5C	130
***Hgi*CII** *Herpetosiphon giganteus Hpg9*	*RM*	G'GWCC	273 437 m^5C	130
***Hgi*DI** *Herpetosiphon giganteus Hpa2*	*RM*	GR'CGYC	359 309 m^5C	129 131
***Hgi*DII** *Herpetosiphon giganteus Hpa2*	*RM*	G'TCGAC	354 m^5C	129
***Hgi*EI** *Herpetosiphon giganteus Hpg24*	*RM*	G'GWCC	274 437 m^5C	130

Type-II R-M systems

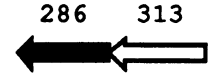
System	Genes	Specificity	Gene organization (R / M)	Refs
HgiGI *Herpetosiphon giganteus* *HpaI*	*RM*	GR'CGYC	 m^5C	129
HhaI *Haemophilus haemolyticus*	*RM*	GCG'C	327  m^5C	132 133
HhaII *Haemophilus haemolyticus*	*RM*	G'ANTC	227 228  $m^6A\beta$	134 135 136
HinlII *Haemophilus influenzae RFL1*	*M*	CATG'		101
HincII[8] *Haemophilus influenzae Rc*	*RM*	GTY'RAC	257 518  $m^6A\gamma$	101 137 138 139
HindII *Haemophilus influenzae Rd*	*M*	GTY'RAC	m^6A	12
HindIII *Haemophilus influenzae Rd*	*RM*	A'AGCTT	300 309  $m^6A\beta$	12 140
HinfI *Haemophilus influenzae Rf*	*RM*	G'ANTC	262 358  $m^6A\beta$	77 141
HinpII *Haemophilus influenzae P1*	*RM*	G'CGC		133
HjaI *Hyphomonas jannaschiana*	*RM*	GAT'ATC		142

[8]The *hincIIM* gene was reported to be 502 codons long (137); independent analysis indicates that it is probably 518 codons (138).

Type-II R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
HpaI *Haemophilus parainfluenzae*	RM	GTT'A$\underline{A}$C	254 314	$m^6A\beta$	143 144
HpaII *Haemophilus parainfluenzae*	RM	C'$\underline{C}$GG	358	m^5C	145 146
KasI *Kluyvera ascorbata*	RM	G'CCGGC			31
KpnI *Klebsiella pneumonia OK8*	M	GGTAC'C		m^6A	147 148
Kpn2I *Klebsiella pneumonia RFL2*	RM	T'CCGGA			87
MluI *Micrococcus luteus*	RM	A'CGCGT			76
MspI *Moraxella sp.*	RM	$\underline{C}$'CGG	262 418	m^5C	16 139 149 150
MstI *Microcoleus sp.*	M	TGC'GCA			81
MunI *Mycoplasma sp.*	RM	C'AATTG	202 229	$m^6A\beta$	144
MvaI *Micrococcus varians RFL19*	RM	C$\underline{C}$'WGG	259 454	$m^4C\alpha$	55 90 151
MwoI *Methano-bacterium wolfei*	RM	GCN_5'N_2GC	256	$m^4C\beta$	152 153

Type-II R-M systems

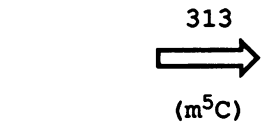
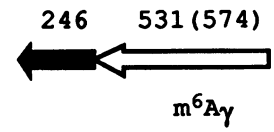
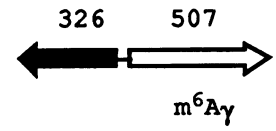
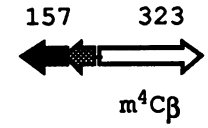
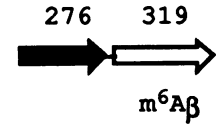
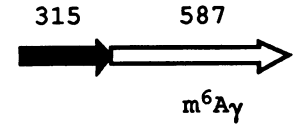
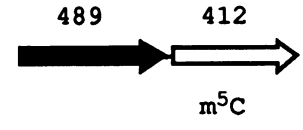
System	Genes	Specificity	Gene organization R M	Refs
***Nae*I** *Nocardia aerocolonigenes*	*M*	GGC'GCC		125
***Nci*I** *Neisseria cinerea*	*RM*	CC'SGG		96
***Nco*I** *Nocardia corallina*	*RM*	C'CATGG	287 422 	52 125
***Nde*I** *Neisseria denitrificans*	*RM*	CA'TAATG	368 478  m^6A_α	87 154 155
***Nde*II** *Neisseria denitrificans*	*M*	'GATC		76
***Ngo*BI** *Neisseria gonorrhoea WR220*	*RM*	GG'CC		156
***Ngo*BIII** *Neisseria gonorrhoea WR220*	*M*	GGN_2CC		156
***Ngo*DI** *Neisseria gonorrhoea 1291*	*M*	GCSGC		139
***Ngo*DII** *Neisseria gonorrhoea 1291*	*M*	RGCGC'Y		139
***Ngo*DIII** *Neisseria gonorrhoea 1291*	*M*	GATC		156
***Ngo*MI** *Neisseria gonorrhoea MS11*	*RM*	G'CCGGC	286 313  m^5C	157

Type-II R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
NgoMII *Neisseria gonorrhoea MS11*	*M*	TCACC			157
NgoMIII *Neisseria gonorrhoea MS11*	*M*	CCGC'GG			157
NgoMIV *Neisseria gonorrhoea MS11*	*M*	GGN$_2$CC		399 → m^5C	139, 158
NgoPI *Neisseria gonorrhoea P9*	*M*	RGCGCY			159
NgoPII[9] *Neisseria gonorrhoea P9*	*RM*	GG'CC	278	341(330) → m^5C	51, 160
NheI *Neisseria mucosa heidelbergensis*	*M*	G'CTAGC			76
NlaI *Neisseria lactamica*	*M*	GGCC			80
NlaIII *Neisseria lactamica*	*RM*	CATG'	230	334 → m^6A_α	12, 161, 162
NlaIV *Neisseria lactamica*	*RM*	GGN'NCC			12, 163
NlaV *Neisseria lactamica*	*M*	CCGG			80

[9]An alternative start for the *ngoPIIM* gene would make it 330 codons long.

Type-II R-M systems

System	Genes	Specificity	Gene organization R / M	Refs
NlaX[10] *Neisseria lactamica*	*(M)*	?		162
PaeR7I[11] *Pseudomonas aeruginosa pMG7*	*RM*	C'TCGAG		164 165 166
PflMI *Pseudomonas fluorescens*	*M*	$CCAN_4$'NTGG		76
PstI *Providencia stuarti*	*RM*	CTGCA'G		167 168
PvuII *Proteus vulgaris*	*RM*	CAG'CTG		60 84 169 170
RsrI *Rhodopseudomonas sphaeroides*	*RM*	G'AATTC		50 171 172
SacII *Streptomyces achromogenes*	*RM*	CCGC'GG		12
SalI *Streptomyces albus*	*RM*	G'TCGAC		173 174
Sau3AI *Staphylococcus aureus 3A*	*RM*	'GATC		87 175

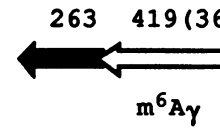
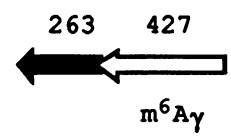
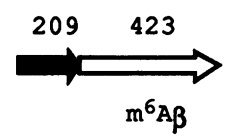
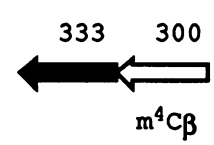
[10]*NlaX* is a putative MTase, the gene for which lies downstream of *nlaIIIM*. The gene was cloned with *nlaIIIM*, and was discovered during sequencing. The *nlaXM* ORF closely resembles an m^5C-MTase. Its specificity is unknown, and it is not clear whether it has a companion endonuclease.

[11]An alternative start for the *paeR7IM* gene would make it 574 codons long.

Type-II R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
Sau96I *Staphylococcus aureus PS96*	*RM*	G'GNCC	261 430	m^5C	140 176
Sbo13I *Shigella boydii 13*	*RM*	TCG'CGA			177
SduI *Streptococcus durans RFL3*	*RM*	GDGCH'C			123
SfiI *Streptomyces fimbriatus*	*RM*	GGCCN$_4$'NGGCC	269 421	$m^4C\beta$	178
SinI *Salmonella infantis*	*RM*	G'GWCC	230 461	m^5C	179
SmaI *Serratia marcescens*	*RM*	CCC'GGG	247 292	$m^4C\alpha$	53 102 180
SpeI *Sphaerotilus natans*	*M*	A'CTAGT			76
SphI *Streptomyces phaeochromogenes*	*M*	GCATG'C	316	$m^6A\alpha$	12
SspI *Sphaerotilus natans*	*M*	AAT'ATT			181
SssI *Spiroplasma species MQ1*	*M*	CG	386	m^5C	182
StyI *Salmonella typhi 27*	*RM*	C'CWWGG			143

Type-II R-M systems

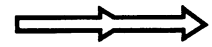
System	Genes	Specificity	Gene organization (R / M)	Refs
TaqI[12] *Thermus aquaticus YT1*	*RM*	T'CGA		18 184 185
ThyI *Treponema Hyodesenteriae*	*M*	CTGCAG		139
TliI *Thermococcus litoralis*	*M*	CTCGAG		76
TthHB8I *Thermus thermophilus HB8*	*RM*	T'CGA		185
XbaI *Xanthomonas badrii*	*RM*	T'CTAGA		125 126
XcmI *Xanthomonas campestris*	*M*	CCAN5'N4TGG		76
XhoI *Xanthomonas holcicola*	*RM*	C'TCGAG		186
XmaI *Xanthomonas malvacaerum*	*RM*	C'CCGGG		54
XmnI *Xanthomonas manihotis*	*RM*	GAAN2'N2TTC		76

[12] The *taqIM* gene was erroneously reported to be 363 codons long. It is now thought to 419 codons (183).

Type IIs R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
AciI *Arthrobacter citreus*	*M*	CCGC			54
Alw26I *Acinetobacter lwoffi RFL26*	*M*	GTCTC 1/5		$m^5C + m^6A$	77
BbvI *Bacillus brevis*	*RM*	GCAGC 8/12		374 → m^5C	31
BcgI *Bacillus coagulans*	*RM*	10/12 $CGAN_6TGC$ 12/10			34
EarI *Enterobacter aerogenes*	*M*	CTCTTC 1/4			54
Eco31I *Escherichia coli RFL31*	*M*	GGTCTC 1/5		$m^5C + m^6A$	77
Eco57I *Escherichia coli RFL57*	*RMM*	CTGAAG 16/14	993 → $R + m^6A\gamma$	544 ← $m^6A\gamma$	123
Esp3I *Erwinia sp.*	*M*	CGTCTC 1/5		$m^5C + m^6A$	77
FokI *Flavobacterium okeanokoites*	*RMM*	GGATG 9/13	578 ←	647 → $m^6A\alpha + m^6A\alpha$	29, 30, 101
GsuI *Gluconobacter suboxydans H-15T*	*M*	CTGGAG 16/14			123
HgaI *Haemophilus gallinarum*	*RMM*	GACGC 5/10	488 ←	358, 357 → m^5C m^5C	31

Type IIs R-M systems

System	Genes	Specificity	Gene organization R	Gene organization M	Refs
***Lla*I**[13] *Lactococcus lactis*	*MM*	?		622  $m^6A\alpha$ + $m^6A\alpha$	187 188
***Mbo*II** *Moraxella bovis*	*RM*	GAAGA 8/7	416 	260 $m^6A\beta$	189
***Sfa*NI** *Streptococcus faecalis ND547*	*MM*	GCATC 5/9			31
***Uba*1109I** *Unidentified bacterium*	*RM*	GCAGC 8/12			87

[13]The specificity of *Lla*I is uncertain. The *M* gene was identified on a restriction-resistant transducing phage. The sequence indicates that the gene encodes a double m^6A-MTase, similar to M·*Fok*I. This suggests that *Lla*I is a type-IIs system.

Type III R-M systems

System	Type	Specificity	Gene organization R	Gene organization M	Refs
EcoP1 *Escherichia coli phage P1*	*RM*	AGAACC	970	646 $m^6A\beta$	9
EcoP15 *Escherichia coli 15T⁻ [p15B]*	*RM*	CAGCAG		645 $m^6A\beta$	9
StyLTI *Salmonella typhimurium LT7*	*RM*	CAGAAG		m^6A	190

Solitary restriction and modification functions

System	Function	Specificity	Gene organization	Refs
BamH2[14] *Bacillus amyloliquefaciens phage H2*	M	GGATCC	265(279) ⇨ $m^4C\beta$	191
CpG MTase[15] *Homo sapiens*	M	CG	m^5C	193
CpG MTase[16] *mouse*	M	CG	1523 ⇨ m^5C	194 195
CviAII[17] *Chlorella virus PBCV1*	M	?	⇨ (m^5C)	104
CviBIII *Chlorella virus NC-1A*	M	TCGA	377 ⇨ $m^6A\gamma$	196
Dam$_{Eco}$ *Escherichia coli*	M	GATC	278 ⇨ $m^6A\alpha$	197 198

[14]This enzyme, the '*Bacillus amyloliquefaciens* proviral H2, *Bam*HI methyltransferase' (191), is distinct from the methyltransferase of the *Bam*HI R-M system (82,86), and from the H2 m^5C-methyltransferase from T.Trautner's group (192). An alternative start for the gene would make it 279 codons in length.

[15]Only the 3'-end of the human CpG MTase gene has been cloned. It encodes the CO_2H-terminus of the MTase, the sequence of which is similar to those of bacterial m^5C-MTases.

[16]The mouse CpG methyltransferase was initially thought to be 1573 aa long, but it is now thought to be 1523 aa . The CO_2H-terminal 550 aa resemble bacterial m5C-MTases (194). The enzyme has a 35-fold preference for hemimethylated substrates, rather than unmethylated substrates, and a 50 to 200-fold specificity towards 5'-CG-3'.

[17]*Cvi*AII is a putative MTase, that was discovered serendipitously during sequencing. The ORF resembles an m^5C-MTase, and is closely similar to another *Chlorella* virus methyltransferase, M·*Cvi*JI . Its specificity is unknown, and it is not clear whether it has a companion endonuclease.

Solitary restriction and modification functions

System	Function	Specificity	Gene organization	Refs
Dam_{Ec67}[18] *Escherichia coli Cl-1*	M	G<u>A</u>TC	285 ⇨ $m^6A\alpha$	199
Dam_{P1} *Escherichia coli phage P1*	M	G<u>A</u>TC	277 ⇨ $m^6A\alpha$	200
Dam_{T1} *Escherichia coli phage T2*	M	G<u>A</u>TC	237 ⇨ $m^6A\alpha$	201
Dam_{T2}[19] *Escherichia coli phage T2*	M	G<u>A</u>TC	259 ⇨ $m^6A\alpha$	204
Dam_{T4} *Escherichia coli phage T4*	M	G<u>A</u>TC	259 ⇨ $m^6A\alpha$	205 206 207
Dcm[20] *Escherichia coli*	M	C<u>C</u>WGG	472 ⇨ m^5C	40 209 210
***Dpn*I**[21] *Streptococcus pneumoniae*	R	$Gm^6A'TC$	171 ➡	108
ϕ3T I *Bacillus subtilis phage ϕ3T*	M	G<u>C</u>NGC GG<u>C</u>C	443 ⇨ m^5C	37 212 213

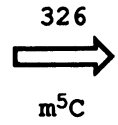
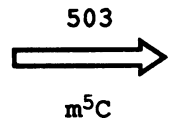
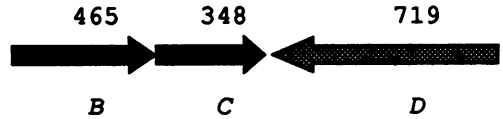
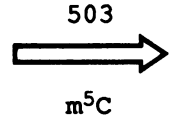
[18] Dam_{Ec67} is encoded by a retroviral-like element, which also codes for a reverse transcriptase/RNase H enzyme (RT). Three Dam sites occur in the promoter region for the RT gene..

[19] Relaxed specificity, hypermethylating mutants (Dam^h) of the T2 and T4 Dam methylases have been isolated. The mutants, Pro^{126} to Ser^{126}, methylate AGACC in addition to GATC (202,203)

[20] Immediately downstream of *dcm* is *vsr*, the gene for very short patch repair (40). Vsr functions at mismatched Dcm sites (208).

[21] *Dpn*I requires methylation of the A residues in order to cleave (211).

Solitary restriction and modification functions

System	Function	Specificity	Gene organization	Refs
ϕ3T II *Bacillus subtilis phage ϕ3T*	M	?TCGA?		214
H2 *Bacillus amyloliquefaciens phage H2*	M	GDGCHC GCNGC GGCC		192
McrA[22] *Escherichia coli*	R	Cm5CGG?		45 217
McrB[23] *Escherichia coli*	R	GmC?		20 45 46 47 219
Mrr[24] *Escherichia coli*	R	m6A? m5CG?		20 220 221
ρ11b *Bacillus subtilis phage ρ11*	M	GCNGC GGCC	m5C	37 212
ρ11s[25] *Bacillus subtilis phage ρ11s*	M	GDGCHC GGCC		222

[22]The specificity of *Mcr*A is uncertain, but it requires the presence of modified cytosine (215,216).

[23]The specificity of *Mcr*B is also uncertain, but it requires the presence of modified cytosine (169,215,216,218). The modification can be m^4C, m^5C, or hm^5C. An alternative start for *mcrC* would make the gene 358 codons long.

[24]The specificity of *Mrr* is also uncertain, but it requires the presence of either modified adenine or cytosine (43,220,221).

[25]Chimeric methyltransferases have been constructed between ρ11s, ϕ3T I and SPR by interchanging equivalent sections of their genes. Some of the recombinants, formed in vitro by crossing over within the sections of the genes that specify the target recognition domains (TRDs), display novel

Solitary restriction and modification functions

System	Function	Specificity	Gene organization	Refs
SPβ *Bacillus subtilis phage SPβ*	M	GCNGC GGCC	443 ⇨ m^5C	37 212 213
SPR *Bacillus subtilis phage SPR*	M	CCWGG CCGG GGCC	439 ⇨ m^5C	100 212 223 224 225

combinations of the parental sequence-specificities. One chimera, between ρ11s and ϕ3T I, methylates GDGCHC, GCNGC and GGCC; another, between ρ11s and SPR, methylates GDGCHC, CCGG and GGCC (58). A third, between ϕ3T I and SPR, methylates GCNGC, CCGG and GGCC (57).

[a] Only systems that have been cloned, at least partially, are listed. See refs. 6 and 7 for surveys of all R-M systems.

[b] Genes known to be completely cloned are listed. *R* signifies that the endonuclease gene has been cloned; *M* signifies that the methyltransferase gene has been cloned; *S* signifies that the specificity gene has been cloned.

[c] Only one strand of the recognition sequence is shown, printed 5′ to 3′. For endonucleases that cut outside of the recognition sequence, the sequence shown is the one that occurs on the 5′ side of the cut. Strings of unspecified nt are designated N_i, where i specifies the number of nt in the string. The standard abbreviations for alternative nts are:

R: A or G	M: A or C	S: C or G	H: A or Y	V: C or R
Y: C or T	K: G or T	W: A or T	B: G or Y	D: T or R

Cleavage positions are indicated by apostrophes, for type II enzymes, and by numerals for types IIs, IV and V enzymes. For the former, only the cut on one strand is shown; for the other systems, the cuts on both strands are shown.

The nts that probably become methylated are indicated by underlining. For type II systems, only the methylated base on one strand is shown. For the other types the methylated bases on both strands are shown. A and C signify that these bases are methylated; T and G signify that the complementary bases are methylated. The methylated base is often inferred from the sensitivity of the DNA to restriction, or from the aa sequence of the methyltransferase, rather than from direct chemical analysis, and so the assignments should be regarded as tentative.

[d] Genes are depicted as arrows; the directions indicate transcriptional orientation. The diagrams are arranged so that the *R* gene is always on the left and *M* gene is on the right. The genes, and the intervals between them, are drawn to scale. Unsequenced genes are shown as thin arrows: →. Sequenced genes are shown as fat arrows; if the sequence is complete, the gene length (in codons) is printed above the arrow; if the sequence is incomplete, no length is given. Endonuclease genes are shown filled: ➡. Methyltransferase genes are open: ⇨. Specificity genes are cross-hatched: ▰. Control genes, and adjacent ORFs, are stippled: ▱.

The probable methylation product is printed beneath the *M* gene. The m4C-MTases and m^6A-MTases are divisible according to aa sequence architecture (55,67). The subclasses are referred to here by subscripts α, β and γ. In the α class, the F-G-G motif (motif 1) occurs before the D/SPPY motif (motif 2). In the β class the order is reversed. In the γ class, the motifs are P-G-G then NPPY. An alternative classification scheme is: 12 D and 12 S (m^6Aα and m^4Cα), 21 D and 21 S (m^6Aβ and m^4Cβ), and N or 12 N (m^6Aγ) (68).

[e] References relate to the cloning or sequencing of the genes, only. See refs. 6, 7 and 69 for references to the discovery and characterization of the enzymes.

DISCUSSION

Restriction and modification enzymes are remarkable for the variety of their specificities. Approximately two hundred different specificities have been discovered so far; probably, many more remain to be found. Why do so many specificities exist? One explanation might be that the enzymes have arisen many times during evolution, and since numerous DNA sequences can be targets for restriction, and no one sequence is universally more appropriate than any other, each specificity represents a slightly different, but equally effective, solution to the same problem. The lack of homology between endonucleases supports the idea that they arose independently; the diverse ways in which *R* and *M* genes are linked supports the idea that the systems assembled independently.

In only one other system—the immune system—does such a variety of specificities occur. The parallel is apt because R-M systems play an equivalent role, in bacteria, as the immune system plays in higher organisms. Whereas immunoglobulins recognize mainly proteins, however, and individuals have a wide repertoire of specificities at their disposal, restriction enzymes recognize DNA, and individuals sport only a few specificities. Our interest in understanding the basis of protein-DNA recognition makes restriction and modification enzymes an attractive group to study. The sequences of many of the proteins are now known; the next step is the determination of their 3-dimensional structures.

ACKNOWLEDGEMENTS

Some of the information summarized here was reported at the DNA methylation meeting in Berlin, 2–7 September 1990. I would like to thank the participants for providing this information prior to its publication. Particular thanks to Drs. Arvydas Janulaitis, Antal Kiss, Manfred Kröger, Noreen Murray, Dan Stein, and Tom Trautner for corrections, for helpful suggestions, and for up-to-the-minute information. Special thanks also to the meeting organizers, Drs. Mario Noyer-Weidner and Tom Trautner, and to their assistants, for making it so enjoyable; and to Dr. Don Comb for generous financial support.

REFERENCES

The nt sequences of some of the genes referred to in this paper are available from GenBank/EMBL. When this is the case, the sequence accession number (e.g. X05242) is printed below, between the year of publication and the name of the journal.

1. Wilson, G.G. (1988) *Gene*, **74**, 281–289.
2. Yuan, R. and Hamilton, D.L. (1984) In Razin, A., Cedar, H. and Riggs, A.D. (eds.), DNA Methylation: Biochemistry and Biological Significance. Springer-Verlag, New York, pp. 11–37.
3. Bickle, T.A. (1987) In Neidhardt, F.C. (ed.), Escherichia coli and Salmonella typhimurium. Amer. Soc. Microbiology, Washington, Vol.1, pp. 692–696.
4. Bennett, S.P. and Halford, S.E. (1989) In Horecker, B.L. (ed.), Current Topics in Cellular Regulation. Academic Press, Vol.30, pp. 57–104.
5. Davies, J.K. (1989) *Clin. Microbiol. Rev.*, **2**, S78–S82.
6. Roberts, R.J. (1990) *Nucleic Acids Res.*, **18**, supplement 2331–2365.
7. Kessler, C. and Manta, V. (1990) *Gene*, **92**, 1–248.
8. Szybalski, W., Kim, S.C., Hasan, N. and Podhajska, A.J. (1991) *Gene*, in press.
9. Hümbelin, M., Suri, B., Rao, D.N., Hornby, D.P., Eberle, H., Pripfl, T., Kenel, S. and Bickle, T.A. (1988) *J. Mol. Biol.*, **200**, 23–29.
10. Xia, Y., Burbank, D.E. and Van Etten, J.L. (1986) *Nucleic Acids Res.*, **14**, 6017–6030.
11. Laue, F., Evans, L.R., Jarsch, M., Brown, N.L. and Kessler, C. (1991) *Gene*, **97**, 87–95.
12. Lunnen, K.D., Barsomian, J.M., Camp, R.R., Card, C.O., Chen, S.Z., Croft, R., Looney, M.C., Meda, M.M., Moran, L.S., Nwankwo, D.O., Slatko, B.E., Van Cott, E.M. and Wilson, G.G. (1988) *Gene*, **74**, 25–32.
13. Borck, K., Beggs, J.D., Brammar, W.J., Hopkins, A.S. and Murray, N.E. (1976) *Mol. Gen. Genet.*, **146**, 199–207.
14. Szomolányi, E., Kiss, A. and Venetianer, P. (1980) *Gene*, **10**, 219–225.
15. Janulaitis, A., Povilionis, P. and Sasnauskas, K. (1982) *Gene*, **20**, 197–204.
16. Walder, R.Y., Langtimm, C.J., Chatterjee, R. and Walder, J.A. (1983) *J. Biol. Chem.*, **258**, 1235–1241.
17. Kiss, A., Pósfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) X02988, *Nucleic Acids Res.*, **13**, 6403–6421.
18. Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., Van Cott, E.M. and Wilson, G.G. (1987) *Nucleic Acids Res.*, **15**, 9781–9796.
19. Yuan, R. (1981) *Ann. Rev. Biochem.*, **50**, 285–315.
20. Sain, B. and Murray, N.E. (1980) *Mol. Gen. Genet.*, **180**, 35–46.
21. Studier, F.W. and Bandyopadhyay, P.K. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4677–4681.
22. Loenen, W.A.M., Daniel, A.S., Braymer, H.D. and Murray, N.E. (1987) X06545, *J. Mol. Biol.*, **198**, 159–170.
23. Daniel, A.S., Fuller-Pace, F.V., Legge, D.M. and Murray, N.E. (1988) *J. Bacteriol.*, **170**, 1775–1782.
24. Price, C., Lingner, J., Bickle, T.A., Firman, K. and Glover, S.W. (1989) X13145; M27781 and M27782, *J. Mol. Biol.*, **205**, 115–125.
25. Smith, H.O. and Kelly, S.V. (1984) In Razin, A., Cedar, H. and Riggs, A.D. (eds.), DNA Methylation: Biochemistry and Biological Significance. Springer-Verlag, New York, pp. 39–71.
26. Wilson, G.G. (1988) *Trends in Genetics*, **4**, 314–318.
27. Landry, D., Looney, M.C., Feehery, G.R., Slatko, B.E., Jack, W.E., Schildkraut, I. and Wilson, G.G. (1989) *Gene*, **77**, 1–10.
28. Sugisaki, H., Kita, K. and Takanami, M. (1989) *J. Biol. Chem.*, **264**, 5757–5761.
29. Kita, K., Kotani, H., Sugisaki, H. and Takanami, M. (1989) J04623; M28828, *J. Biol. Chem.*, **264**, 5751–5756.
30. Looney, M.C., Moran, L.S., Jack, W.E., Feehery, G.R., Benner, J.S., Slatko, B.E. and Wilson, G.G. (1989) M28828, *Gene*, **80**, 193–208.
31. Barsomian, J., Landry, D., Moran, L., Slatko, B., Feehery, G., Jack, W., Nwankwo, D. and Wilson, G. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 11.
32. Bitinaite, J., Maneliene, Z., Grigaite, R., Menkevicius, S., Butkus, V. and Janulaitis, A. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 128.
33. Petrusyte, M., Bitiniaite, J., Menkevicius, S., Klimasauskas, S., Butkus, V. and Janulaitis, A. (1988) *Gene*, **74**, 89–91.
34. Kong, H.M., Morgan, R., Jack, W.E., Nwankwo, D.O. and Schildkraut, I., personal communication.
35. Kramarov, V.M., Fomenkov, A.I. and Matvienko, N.I. (1988) *Bioorg Khim*, **14**, 916–20.
36. Krüger, D.H., Bickle, T.A., Reuter, M., Pein, C.-D. and Schroeder, C. (1990) In Clawson, G.A. (ed.), Nucleic Acid Methylation. Alan R. Liss, New York, pp. 113–124.
37. Noyer-Weidner, M., Jentsch, S., Kupsch, J., Bergbauer, M. and Trautner, T.A. (1985) *Gene*, **35**, 143–150.
38. Vanek, P.G., Connaughton, J.E., Kaloss, W.D. and Chirikjian, J.G. (1990) *FASEB J.*, **4**, A2295.
39. Barras, F. and Marinus, M.G. (1989) *Trends in Genetics*, **5**, 139–143.
40. Sohail, A., Lieb, M., Dar, M. and Bhagwat, A.S. (1990) *J. Bacteriol.*, **172**, 4214–4221.
41. Lacks, S. and Greenberg, B. (1977) *J. Mol. Biol.*, **114**, 153–168.
42. Sladek, T.L. and Maniloff, J. (1987) *Isr J Med Sci*, **23**, 423–6.
43. Heitman, J. and Model, P. (1987) *J. Bacteriol.*, **169**, 3243–3250.
44. MacNeil, D.J. (1988) *J. Bacteriol.*, **170**, 5607–5612.
45. Raleigh, E.A., Trimarchi, R. and Revel, H. (1989) *Genetics*, **122**, 279–296.
46. Ross, T.K., Achberger, E.C. and Braymer, H.D. (1989) *J. Bacteriol.*, **171**, 1974–1981.
47. Noyer-Weidner, M., Krüger, T., Grund, C. and Wild, C. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 45.
48. Macaluso, A. and Mettus, A.M. (1991) *J. Bacteriol.*, **173**, 1353–1356.
49. Chandrasegaran, S. and Smith, H.O. (1988) In Sarma, M.H. and Sarma, R.H. (eds.), Structure and Expression. Adenine Press, Guilderland, NY, Vol.1, pp. 149–156.
50. Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1989) X14697, *Gene*, **85**, 1–13.
51. Sullivan, K.M. and Saunders, J.R. (1989) *Mol. Gen. Genet.*, **216**, 380–387.

52. Zhang, B., Van Cott, E.M., Benner, J.S., Moran, L.S., Slatko, B.E. and Wilson, G.G., (unpublished).
53. Heidmann, S., Seifert, W., Kessler, C. and Domdey, H. (1989) X16458, *Nucleic Acids Res.*, **17**, 9783–9796.
54. Lunnen, K.D. and Wilson, G.G., (unpublished).
55. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkienė, D., Butkus, V. and Janulaitis, A. (1989) X16985 and X17022, *Nucleic Acids Res.*, **17**, 9823–9832.
56. Pósfai, J., Bhagwat, A.S., Pósfai, G. and Roberts, R.J. (1989) *Nucleic Acids Res.*, **17**, 2421–2435.
57. Balganesh, T.S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K. and Trautner, T.A. (1987) *EMBO J.*, **6**, 3543–3549.
58. Trautner, T.A., Balganesh, T.S. and Pawlek, B. (1988) *Nucleic Acids Res.*, **16**, 6649–6658.
59. Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) *J. Mol. Biol.*, **206**, 305–312.
60. Tao, T., Walter, J., Brennan, K.J., Cotterman, M.M. and Blumenthal, R.M. (1989) X13778, *Nucleic Acids Res.*, **17**, 4161–4175.
61. Lauster, R. (1989) *J. Mol. Biol.*, **206**, 313–321.
62. Gough, J.A. and Murray, N.E. (1983) J01630, J01631 and J01632, *J. Mol. Biol.*, **166**, 1–19.
63. Fuller-Pace, F.V. and Murray, N.E. (1986) M14984 and K03551, *Proc. Natl. Acad. Sci. USA*, **83**, 9368–9372.
64. Cowan, G.M., Gann, A.A.F. and Murray, N.E. (1989) J03150 and J03162, *Cell*, **56**, 103–109.
65. Kannan, P., Cowan, G.M., Daniel, A.S., Gann, A.A.F. and Murray, N.E. (1989) *J. Mol. Biol.*, **209**, 335–344.
66. Gann, A.A.F., Campbell, A.J.B., Collins, J.F., Coulson, A.F.W. and Murray, N.E. (1987) Y00524, *Mol. Microbiol.*, **1**, 13–22.
67. Klimasauskas, S., Steponaviciene, D., Maneliene, Z., Petrusyte, M., Butkus, V. and Janulaitis, A. (1990) *Nucleic Acids Res.*, **18**, 6607–6609.
68. Timinskas, A. and Janulaitis, A., personal communication.
69. McClelland, M. and Nelson, M. (1988) *Gene*, **74**, 291–304.
70. Fuller-Pace, F.V., Cowan, G.M. and Murray, N.E. (1985) *J. Mol. Biol.*, **186**, 65–75.
71. Daniel, A.S., Cowan, G. and Murray, N.E., personal communication.
72. Kelleher, J., Daniel, A.S. and Murray, N.E., personal communication.
73. Skrzypek, E. and Piekarowicz, A. (1989) *Plasmid*, **21**, 195–204.
74. Fuller-Pace, F.V., Bullas, L.R., Delius, H. and Murray, N.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6095–6099.
75. Nagaraja, V., Shepherd, J.C.W. and Bickle, T.A. (1985) *Nature*, **316**, 371–372.
76. Nwankwo, D.O., personal communication.
77. Mitkaite, G., Vaisvila, R. and Janulaitis, A., personal communication.
78. Tao, T., Zhang, B.-H., Blumenthal, R. and Wilson, G.G., (unpublished).
79. Karreman, C. and de Waard, A. (1990) M28051, *J. Bacteriol.*, **172**, 266–272.
80. Morgan, R.D., personal communication.
81. Meda, M.M., Chen, S.Z., Perler, F. and Wilson, G.G., (unpublished).
82. Brooks, J.E., Nathan, P.D., Landry, D., Sznyter, L.A., Waite-Rees, P.A., Ives, C.L., Moran, L.S., Slatko, B.E. and Benner, J.S. (1991) X55285, *Nucleic Acids Res.*, **19**, 841–850.
83. Nathan, P.D., Ives, C.L. and Brooks, J.E., ms. in preparation.
84. Tao, T., Bourne, J.C. and Blumenthal, R.M. (1991) *J. Bacteriol.*, **173**, 1367–1375.
85. Brooks, J.E., Benner, J.S., Heiter, D.F., Silber, K.R., Sznyter, L.A., Jager-Quinton, T., Moran, L.S., Slatko, B.E., Wilson, G.G. and Nwankwo, D.O. (1989) *Nucleic Acids Res.*, **17**, 979–997.
86. Vanek, P.G., Connaughton, J.F., Kaloss, W.D. and Chirikjian, J.G. (1990) X54303, *Nucleic Acids Res.*, **18**, 6145.
87. Lubys, A. and Janulaitis, A., personal communication.
88. Maekawa, Y., Yasukawa, H. and Kawakami, B. (1990) *J. Biochem.*, **107**, 645–649.
89. Povilionis, P., Lubys, A. and Janulaitis, A. (1988) *Genetika (Moscow)*, **24**, 210–215.
90. Menkevicius, S., Butkus, V. and Janulaitis, A., personal communication.
91. Kupper, D., Zhou, J.-G., Venetianer, P. and Kiss, A. (1989) X13555, *Nucleic Acids Res.*, **17**, 1077–1088.
92. Heiter, D., Nathan, P., Benner, J., Marcus, A. and Brooks, J.E., personal communication.
93. Pósfai, G., Kiss, A., Erdei, S., Pósfai, J. and Venetianer, P. (1983) X15758, *J. Mol. Biol.*, **170**, 597–610.
94. Davenis, A., Lubys, A. and Janulaitis, A., personal communication.
95. Vásquez, C., Saavedra, C. and González, E., personal communication.
96. Van Cott, E.M., Charles, J., Sears, L. and Wilson, G.G., (unpublished).
97. Xu, G., Kapfer, W., Walter, J. and Trautner, T.A. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 85.
98. Walter, J., Noyer-Weidner, M. and Trautner, T.A. (1990) X51515, *EMBO J.*, **9**, 1007–1013.
99. Kapfer, W., Walter, J. and Trautner, T.A. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 93.
100. Kiss, A. and Baldauf, F. (1983) *Gene*, **21**, 111–119.
101. Stankevicius, K., Lubys, A. and Janulaitis, A., personal communication.
102. Povilonis, P. and Janulaitis, A., personal communication.
103. Shields, S.L., Burbank, D.E., Grabherr, R. and Van Etten, J.L. (1990) *Virology*, **176**, 16–24.
104. Van Etten, J., personal communication.
105. Stefan, C., Xia, Y. and Van Etten, J.L. (1991) *Nucleic Acids Res.*, **19**, 307–311.
106. Howard, K.A., Card, C., Benner, J.S., Callahan, H.L., Maunus, R., Silber, K., Wilson, G. and Brooks, J.E. (1986) *Nucleic Acids Res.*, **14**, 7939–7951.
107. Sznyter, L.A., Slatko, B., Moran, L., O'Donnell, K.H. and Brooks, J.E. (1987) Y00449, *Nucleic Acids Res.*, **15**, 8249–8266.
108. Lacks, S.A., Mannarelli, B.M., Springhorn, S.S. and Greenberg, B. (1986) *Cell*, **46**, 993–1000.
109. Cerritelli, S., Springhorn, S.S. and Lacks, S.A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9223–9227.
110. Lacks, S.A. and Springhorn, S.S. (1984) *J. Bacteriol.*, **157**, 934–936.
111. Mannarelli, B.M., Balganesh, T.S., Greenberg, B., Springhorn, S.S. and Lacks, S.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4468–4472.
112. Benner, J.S., personal communication.
113. Sznyter, L.A. and Brooks, J.E. (1988) *Gene*, **74**, 53.
114. Sznyter, L., Moran, L., Slatko, B.E. and Brooks, J.E., personal communication.
115. Brenner, V., Venetianer, P. and Kiss, A. (1990) X17111, *Nucleic Acids Res.*, **18**, 355–359.
116. Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E. and Rosenberg, J.M. (1981) *J. Biol. Chem.*, **256**, 2143–2153.
117. Newman, A.K., Rubin, R.A., Kim, S.-H. and Modrich, P. (1981) *J. Biol. Chem.*, **256**, 2131–2139.
118. Kossykh, V.G., Buryanov, Y.I. and Bayev, A.A. (1980) *Mol. Gen. Genet.*, **178**, 717–718.
119. Som, S., Bhagwat, A.S. and Friedman, S. (1987) X05050, *Nucleic Acids Res.*, **15**, 313–332.
120. Kossykh, V., Repyk, A., Kaliman, A. and Buryanov, Y. (1989) X16025, *Biochim. Biophys. Acta*, **1009**, 290–292.
121. Bhagwat, A.S., Johnson, B., Weule, K. and Roberts, R.J. (1990) *J. Biol. Chem.*, **265**, 767–773.
122. Bougueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) M19941, *Nucleic Acids Res.*, **12**, 3659–3676.
123. Vaisvila, R. and Janulaitis, A., personal communication.
124. Malewskij, T., Povilonis, P. and Janulaitis, A. (1991).
125. Van Cott, E.M. and Wilson, G.G. (1988) *Gene*, **74**, 55–59.
126. Zhang, B.-H., Van Cott, E.M. and Wilson, G.G., (unpublished).
127. Slatko, B.E., Croft, R., Moran, L.S. and Wilson, G.G. (1988) *Gene*, **74**, 45–50.
128. Zhang, B.-H., Van Cott, E.M., Lunnen, K.D., Meda, M.M., Benner, J.S. and Wilson, G.G., (unpublished).
129. Düsterhöft, A. (1990) Ph.D. Thesis. JLU, Giessen, W.Germany.
130. Erdmann, D. (1991) Ph. D. Thesis. JLU, Giessen, W. Germany.
131. Düsterhöft, A., Erdmann, D. and Kröger, M. (1991) X55140, *Nucleic Acids Res.*, **19**, 1049–1056.
132. Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G.G. and Wells, R.D. (1987) J02677, *J. Biol. Chem.*, **262**, 4770–4777.
133. Barsomian, J.M., Card, C.O. and Wilson, G.G. (1988) *Gene*, **74**, 5–7.
134. Mann, M.B., Rao, R.N. and Smith, H.O. (1978) *Gene*, **3**, 97–112.
135. Schoner, B., Kelley, S. and Smith, H.O. (1983) K00508, *Gene*, **24**, 227–236.
136. Chandrasegaran, S., Wu, L.P., Valda, E. and Smith, H.O. (1988) *Gene*, **74**, 15–21.
137. Ito, H., Sadaoka, A., Kotani, H., Hiraoka, N. and Nakamura, T. (1990) X52124, *Nucleic Acids Res.*, **18**, 3903–3911.
138. Waite-Rees, P.A., Nwankwo, D.O., Moran, L.S., Slatko, B.E., Wilson, G.G. and Benner, J.S. (1991) X54197, *Nucleic Acids Res.*, in press.
139. Piekarowicz, A., Yuan, R. and Stein, D.C. (1991) *Nucleic Acids Res.*, in press.
140. Nwankwo, D.O., Moran, L.S., Slatko, B.E. and Wilson, G.G. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 76.

141. Chandrasegaran, S., Lunnen, K.D., Smith, H.O. and Wilson, G.G. (1988) *Gene*, **70**, 387–392.
142. Danaher, R.J. and Stein, D.C. (1990) *Gene*, **89**, 129–133.
143. Waite-Rees, P.A., Moran, L.S., Slatko, B.E., Hornstra-Coe, L. and Benner, J.S. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 86.
144. Zareckaja, N., Vaisvila, R. and Janulaitis, A., personal communication.
145. Card, C.O., Wilson, G.G., Weule, K., Hasapes, J., Kiss, A. and Roberts, R.J. (1990) X51322, *Nucleic Acids Res.*, **18**, 1377–1383.
146. Kulakauskas, S. and Janulaitis, A., personal communication.
147. Hammond, A.W., Gerard, G.F. and Chatterjee, D.K. (1990) *Gene*, **97**, 97–102.
148. Kiss, A., Finta, C. and Venetianer, P., personal communication.
149. Nwankwo, D.O. and Wilson, G.G. (1988) *Gene*, **64**, 1–8.
150. Linn, P.M., Lee, C.H. and Roberts, R.J. (1989) X14191, *Nucleic Acids Res.*, **17**, 3001–3011.
151. Povilonis, P., Menkevicius, S., Butkiene, D., Vaisvila, R., Butkus, V. and Janulaitis, A., personal communication.
152. Lunnen, K.D., Morgan, R.D., Timan, C.J., Krzycki, J.A., Reeve, J.N. and Wilson, G.G. (1989) *Gene*, **77**, 11–19.
153. Lunnen, K.D., Moran, L.S., Slatko, B.E. and Wilson, G.G., (unpublished).
154. Silber, K.R., Polisson, C., Rees, P.A. and Benner, J.S. (1988) *Gene*, **74**, 43–44.
155. Waite-Rees, P.A., Polisson, C., Silber, K.R., Moran, L.S., Slatko, B.E. and Benner, J.S., *Nucleic Acids Res.*, submitted for publication.
156. Piekarowicz, A. and Stein, D., personal communication.
157. Chien, R. and Stein, D., personal communication.
158. Stein, D., personal communication.
159. Sullivan, K.M. and Saunders, J.R. (1988) In Poolman, J.T., Zanen, H.C., Meyer, T.F., Heckels, J.E., Måkelå, P.R.H., Smith, H. and Beuvery, E.C. (eds.), Gonococci and Meningococci. Kluwer Academic Publishers, Dordecht, Netherlands, pp. 329–334.
160. Sullivan, K.M. and Saunders, J.R. (1988) X06965, *Nucleic Acids Res.*, **16**, 4369–4387.
161. Barsomian, J., Van Cott, E., Lunnen, K., Meda, M., Moran, L., Morgan, R., Slatko, B., Zhang, B. and Wilson, G. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 105.
162. Labbé, D., Höltke, H.J. and Lau, P.C.K. (1990) X54485, *Mol. Gen. Genet.*, **224**, 101–110.
163. Morgan, R.D., Camp, R.R. and Wilson, G.G., (unpublished).
164. Theriault, G. and Roy, P.H. (1982) *Gene*, **19**, 355–359.
165. Gingeras, T.R. and Brooks, J.E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 402–406.
166. Theriault, G., Roy, P.H., Howard, K.A., Benner, J.S., Brooks, J.E., Waters, A.F. and Gingeras, T.R. (1985) X03274, *Nucleic Acids Res.*, **13**, 8441–8461.
167. Walder, R.Y., Hartley, J.L., Donelson, J.E. and Walder, J.A. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1503–1507.
168. Walder, R.Y., Walder, J.A. and Donelson, J.E. (1984) K02081, *J. Biol. Chem.*, **259**, 8015–8026.
169. Blumenthal, R.M., Gregory, S.A. and Cooperider, J.S. (1985) X13778, *J. Bacteriol.*, **164**, 501–509.
170. Athanasiadis, A., Gregoriu, M., Thanos, D., Kokkinidis, M. and Papamatheakis, J. (1990) X52681, *Nucleic Acids Res.*, **18**, 6434.
171. Kaszubska, W., Aiken, C., O'Connor, C.D. and Gumport, R.I. (1989) X16456, *Nucleic Acids Res.*, **17**, 10403–10425.
172. Stephenson, F.H. and Greene, P.J. (1989) X16456, *Nucleic Acids Res.*, **17**, 10503.
173. Rodicio, M.R. and Chater, K.F. (1988) *Gene*, **74**, 39–42.
174. Alvarez, M.A., Chater, K.F. and Rodicio, M.R. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 91.
175. Seeber, S., Kessler, C. and Götz, F. (1990) M32470, *Gene*, **94**, 37–43.
176. Szilák, L., Venetianer, P. and Kiss, A. (1990) X53096, *Nucleic Acids Res.*, **18**, 4659–4664.
177. Levin, P., Lee, T. and Benner, J.S., personal communication.
178. Van Cott, E.M., Moran, L.S., Slatko, B.E. and Wilson, G.G., (unpublished).
179. Karreman, C. and de Waard, A. (1988) J03391, *J. Bacteriol.*, **170**, 2527–2532.
180. Dunbar, J., personal communication.
181. Eldridge, A., Hornstra, L., Waite-Rees, P. and Benner, J., personal communication.
182. Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G.G., Rottem, S. and Razin, A. (1990) X17195, *Nucleic Acids Res.*, **18**, 1145–1152.
183. Barany, F., personal communication.
184. Barany, F. (1987) *Gene*, **56**, 13–27.
185. Zebala, J.A., Mayer, A.A. and Barany, F. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 16.
186. Brooks, J.E., personal communication.
187. Hill, C., Miller, L. and Klaenhammer, T. (1990) *Plasmid*, **23**, 167.
188. Hill, C. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 98.
189. Bocklage, H., Heeger, K. and Müller-Hill, B. (1991) X56977, *Nucleic Acids Res.*, **19**, 1007–1013.
190. De Backer, O. and Colson, C. (1991) *J. Bacteriol*, **173**, 1321–1327.
191. Connaughton, J.F., Kaloss, W.D., Vanek, P.G., Nardone, G.A. and Chirikjian, J.G. (1990) X53032, *Nucleic Acids Res.*, **18**, 4002.
192. Lange, C., Noyer-Weidner, M., Trautner, T.A., Weiner, M. and Zahler, S.A. (1991) *Gene*, in press.
193. Andrews, P.D., Taylor, C. and Hornby, D.P. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990.
194. Bestor, T., Page, A.W. and Carlson, L.L. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 47.
195. Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V. (1988) X14805, *J. Mol. Biol.*, **203**, 971–983.
196. Narva, K.E., Wendell, D.L., Skrdla, M.P. and Van Etten, J.L. (1987) X06618, *Nucleic Acids Res.*, **15**, 9807–9823.
197. Herman, G.E. and Modrich, P. (1982) *J. Biol. Chem.*, **257**, 2605–2612.
198. Brooks, J.E., Blumenthal, R.M. and Gingeras, T.R. (1983) J01600, *Nucleic Acids Res.*, **11**, 837–851.
199. Hsu, M.Y., Inouye, M. and Inouye, S. (1990) M55249, *Proc Natl Acad Sci USA*, **87**, 9454–9458.
200. Coulby, J.M. and Sternberg, N.L. (1988) *Gene*, **74**, 191.
201. Schneider-Scherzer, E., Auer, B., de Groot, E.J. and Schweiger, M. (1990) J05393, *J. Biol. Chem.*, **265**, 6086–6091.
202. Miner, Z., Schlagman, S. and Hattman, S. (1988) *Gene*, **74**, 275–276.
203. Miner, Z., Schlagman, S.L. and Hattman, S. (1989) *Nucleic Acids Res.*, **17**, 8149–8157.
204. Miner, Z. and Hattman, S. (1988) M22342, *J. Bacteriol.*, **170**, 5177–5184.
205. Schlagman, S.L. and Hattman, S. (1983) *Gene*, **22**, 139–156.
206. Macdonald, P.M. and Mosig, G. (1984) K03113, *EMBO J.*, **3**, 2863–2871.
207. Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., Macdonald, P.M. and Mosig, G. (1985) *J. Bacteriol.*, **164**, 932–937.
208. Lieb, M. (1985) *Mol. Gen. Genet.*, **191**, 118–125.
209. Bhagwat, A.S., Sohail, A. and Roberts, R.J. (1986) *J. Bacteriol.*, **166**, 751–755.
210. Hanck, T., Gerwin, N. and Fritz, H.-J. (1989) X13330, *Nucleic Acids Res.*, **17**, 5844.
211. Lacks, S. and Greenberg, B. (1975) *J. Biol. Chem.*, **250**, 4060–4066.
212. Günthert, U., Reiners, L. and Lauster, R. (1986) *Gene*, **41**, 261–270.
213. Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. and Trautner, T.A. (1986) *Gene*, **42**, 89–96.
214. Terschüren, P.A. and Trautner, T.A. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 96.
215. Revel, H.R. (1983) In Mathews, C.K., Kutter, E.M., Mosig, G. and Berget, P. (eds.), Bacteriophage T4. American Society for Microbiology, Washington, D. C., pp. 156–165.
216. Raleigh, E.A. and Wilson, G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9070–9074.
217. Ramalingam, R., Prasad, R., Raja, M.C. and Dharmalingam, K. Abstract, Second NEB workshop on biological DNA modification, Berlin, 2–7 September 1990, pp. 43.
218. Noyer-Weidner, M., Diaz, R. and Reiners, L. (1986) *Mol. Gen. Genet.*, **205**, 469–475.
219. Dila, D., Sutherland, E., Moran, L., Slatko, B. and , E.R. (1990) M34235, *J. Bacteriol.*, **172**, 4888–4900.
220. Waite-Rees, P.A., Keating, C.J., Moran, L.S., Slatko, B.E., Hornstra, L.J. and Benner, J.S. (1991) *J. Bacteriol.*, in press.
221. Kelleher, J.E. and Raleigh, E.A. (1991) *J. Bacteriol.*, in press.
222. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesh, T.S. and Trautner, T.A. (1987) X05242, *EMBO J.*, **6**, 1137–1142.
223. Behrens, B., Pawlek, B., Morelli, G. and Trautner, T.A. (1983) *Mol. Gen. Genet.*, **189**, 10–16.
224. Buhk, H.-J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Günthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) K02124, *Gene*, **29**, 51–61.
225. Pósfai, G., Baldauf, F., Erdei, S., Pósfai, J., Venetianer, P. and Kiss, A. (1984) K02124, *Nucleic Acids Res.*, **12**, 9039–9049.